United States
Department of
Agriculture

Forest Service

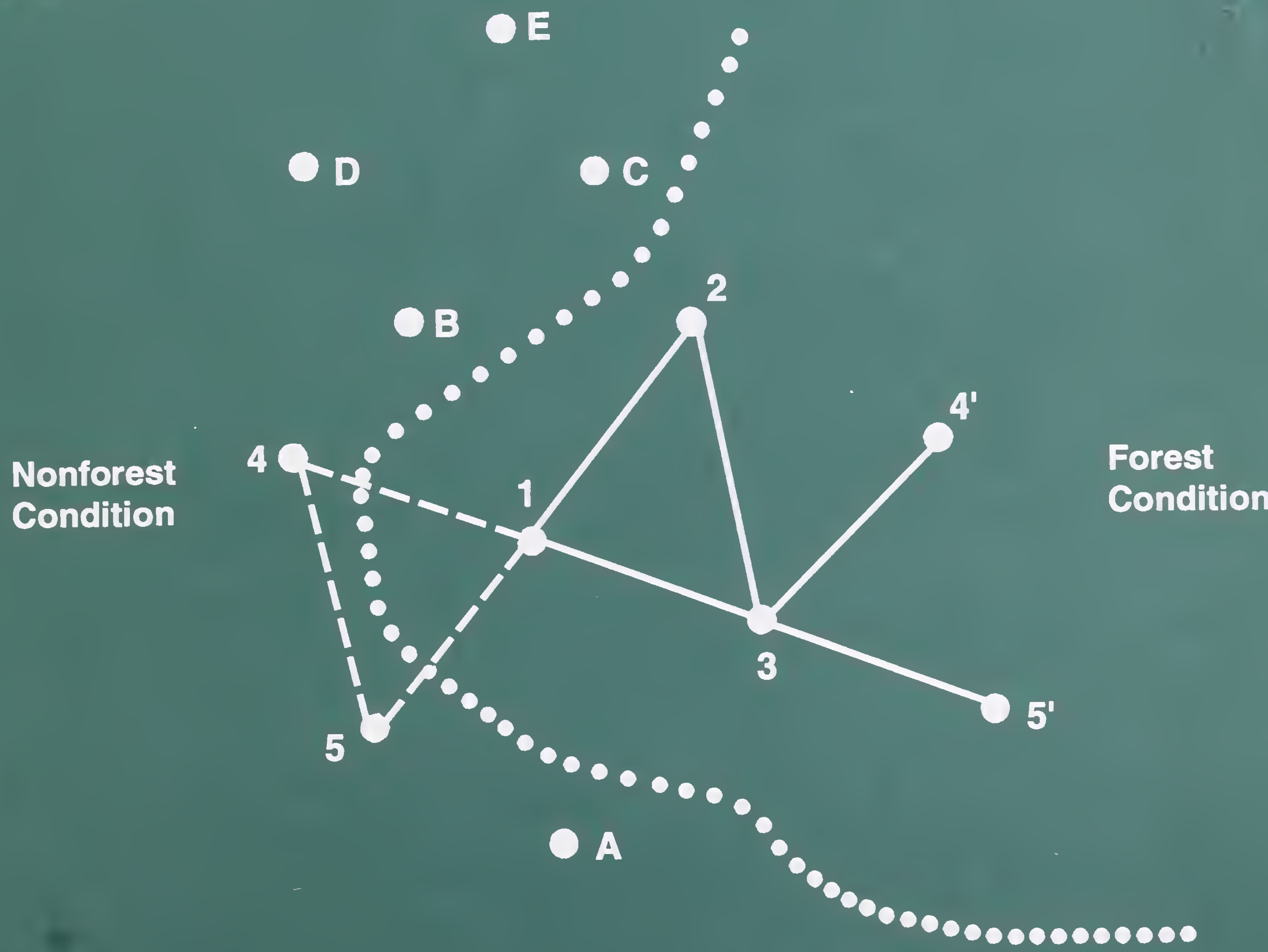
**Rocky Mountain
Forest and Range
Experiment Station**

Fort Collins,
Colorado 80526

**Research Paper
RM-RP-327**

The Extent of Bias Caused by Substituting Points in Forest Survey
A Simulation Study

Michael S. Williams
Hans T. Schreuder
Robin M. Reich



Williams, Michael S., Schreuder, Hans T., Reich, Robin, M. The Extent of Bias Caused by Substituting Points in Forest Survey: A Simulation Study. Res. Pap. RM-RP-327. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 7 p.

Abstract: The bias and efficiency of Forest Inventory and Analysis five-point cluster designs where plots are substituted was compared to five-point cluster designs where plots are not substituted. A simulation study was used to perform the comparison. For populations where trees along the boundary were similar to interior trees, the substituted design gave less biased estimates of basal area than the nonsubstituted designs. Populations with boundary trees that were smaller or larger than interior trees showed a linear increase in bias as the difference between boundary and interior trees increased. For an artificial 67 ha population, biases ranged from -6.2 to +5.1 percent for estimating basal area. In all cases studied, the nonsubstituted design was more efficient than the substituted design. On average, the simulation standard deviation for the nonsubstituted design was about 5 to 6 percent less than for the substituted design.

Keywords: bias, efficiency, five-point cluster design, Forest Inventory and Analysis (FIA), substituted plots, nonsubstituted plots, edge effect, simulation

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please send the publication title and Research Paper RM-RP-327.

Telephone	(490) 498-1719
DG message	D.Schneider:S28A
FAX	(490) 498-1660
E-mail	/s=r.schneider/ou1=s28a@mhs-fswa.attmail.com
Mailing Address	Publications Distribution Rocky Mountain Forest and Range Experiment Station 3825 E. Mulberry Street Fort Collins, CO 80524

USDA Forest Service
Research Paper RM-RP-327

November 1996

The Extent of Bias Caused by Substituting Points in Forest Survey: A Simulation Study

Michael S. Williams
Multiresource Inventory Techniques,
Rocky Mountain Forest and Range Experiment Station[1]

Hans T. Schreuder
Multiresource Inventory Techniques,
Rocky Mountain Forest and Range Experiment Station[1]

Robin M. Reich
Associate Professor, Department of Forest Sciences,
Colorado State University, Fort Collins, Colorado

[1] *Headquarters is in Fort Collins, Colorado, in cooperation with Colorado State University*

Contents

The Extent of Bias Caused by Substituting Points in Forest Survey: A Simulation Study

by Michael S. Williams, Hans T. Schreuder, and Robin M. Reich

INTRODUCTION

The USDA Forest Service, Forest Inventory and Analysis (FIA) units inventory timber resources of the United States approximately every 10 years. The units have considerable freedom in how they obtain the required information in a statistically valid manner and this, in conjunction with differences in forest attributes, leads to interesting differences in approach. All 6 FIA units in the lower 48 states use a 5-point or 10-point cluster in ground sampling, but sampling designs are changing for many of the FIA units. The forest conditions on a point designated as point 1 determine the cluster's condition class. The condition class is determined by the broad forest type, size class, and stand density. Spada (1961) originally recommended a 10-point cluster with an average of 2 sample trees/point. Three of the FIA units use a 5-point cluster and shift points so that all points fall in the same condition class, whereas the other three sample the points regardless of whether they fall in different condition classes. Substituting points into the condition class of sample point 1 has a long history (Lentz 1932, p.17) that has been seriously questioned (L. R. Grosenbaugh, personal communication), but such criticism has been ignored until recently.

This paper documents what the magnitude of estimation bias can be and how estimation efficiency is affected by substituting points. This is done through simulation studies on a mapped population and a number of modifications thereof. We chose the substituted design that would probably be the least biased of the three used.

SAMPLING DESIGN

The following design was used by the Intermountain FIA unit (USDA Forest Service Intermountain Forest Experiment Station 1990). Since point 1 is at the center of the 5-point cluster and point 1 is never moved, the bias is expected to be less than that for the other 2 substituted point designs where point 1 is a corner point of the 5-point cluster that can be shifted from its original position. The additional bias in designs, where point 1 is not the center point of the cluster, is caused by the unequal sampling intensities along the population boundary. When point 1 is the center of the cluster, sampling intensities remain lower along the boundaries, but this reduction in sampling intensity is consistent along all boundaries. This is not case for the other 5-point cluster designs. For the Intermountain design, a point location is systematically established on a map of the population that becomes point 1 of a 5-point cluster. The initial grid pattern of sample points is a "bow-tie" cluster where point 1 is the center point of the cluster and points 2 through 5 surround point 1. Points 1 through 5 are referred to as regular points. This spacing and orientation results in 2 equilateral triangles with 100-ft sides (figure 1). Distance correction for slope is made when necessary.

When the location of all 5 points has been established, substitute points are used when:

1. Any of points 2 through 5 fall in a forest condition that is different from the condition of point 1 and at least 1 acre in size and 120 ft wide. Microsites and inclusions within a condition, and those smaller than 1 acre or less than 120 ft wide, are not considered different from the surrounding condition. These microsites and inclusions are combined with the surrounding condition. Exceptions to the minimum size criteria are improved roads, power-line and pipeline rights-of-way, and operating rail lines. These exceptions are of any area or width.
2. Any of points 2 through 5 fall in an ownership class or geographic area not being sampled (e.g.,

outside the state being inventoried) they are substituted with points from the area originally designated to be sampled.

The following procedures are used for locating substitution points.

1. Figure 2 shows the locations of substitute points A through E. If one of the points 2 through 5 require substitution, point A is used as the first or highest priority substitute point location. Point E is the lowest priority substitute point. The lowest numbered point of locations 2 through 5 requiring substitution is placed in the highest available priority substitute location in the same condition as point 1. This procedure is repeated, beginning with the lowest point requiring substitution, if more than one substitution point is required. Figure 3 shows an example where points 4 and 5 require substitution because they are located in a condition class (clearcut) that differs from the condition class at point 1 (uncut). Although point A is the highest priority point to consider for substitution, it is located in the cut condition class and does not qualify as an appropriate substitute point. Therefore, point 4 is substituted by point B and point 5 by point C.

2. If substitute locations A through E can not be used to locate substitute points, rotation about the regular points is implemented using the following procedure.

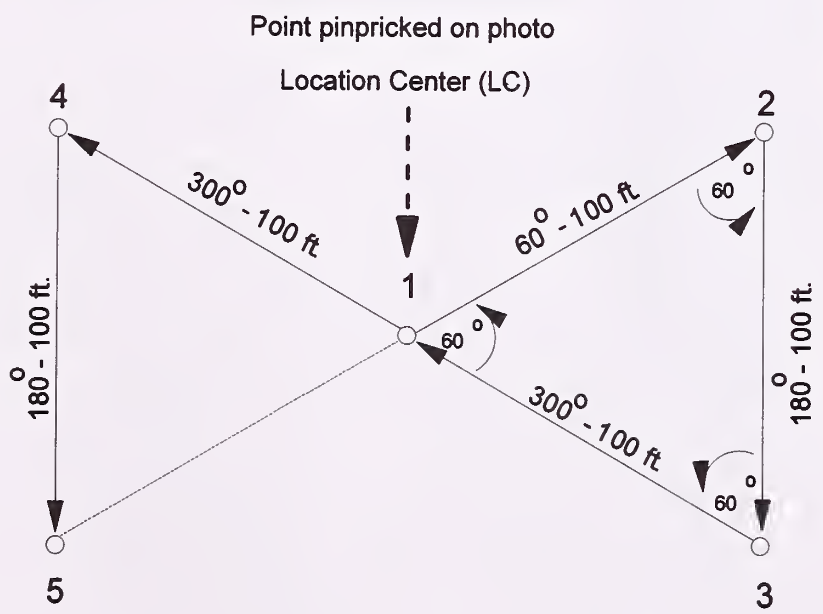


Figure 1. Five-point layout for the variable-radius plot locations.

a) Starting at 0° azimuth from the highest numbered regular point, locate substitute points clockwise at intervals of 60° (and 100 ft) to find the first available qualifying point in the same condition class as point 1 (forming an additional equilateral triangle of points).

b) When more than 1 substitute point is required, substitution of the lowest numbered points is first, continuing the rotation at 60° intervals, selecting other qualifying points forming additional triangles.

c) If necessary, this is repeated at the next highest numbered regular point.

d) When rotation about the regular points does not supply enough substitute points with the same condition class as point 1, rotation is

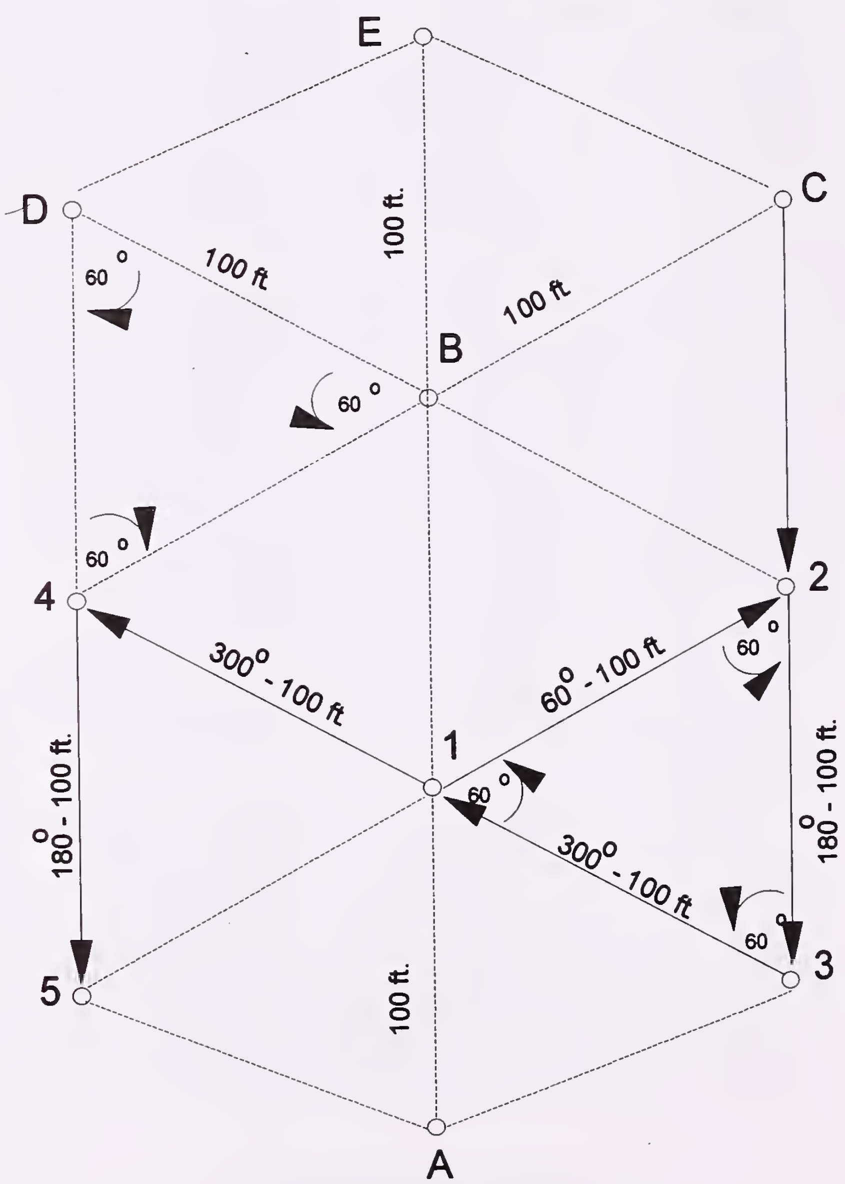


Figure 2. Cluster layout with substitute locations A through E.

performed about each of the previously selected substitute points beginning with the lowest numbered substitute point. For example, in figure 4 points A through E do not qualify as appropriate substitute points for points 4 and 5 because they are not located in the same condition class as point 1 (point 1 is in a forested condition and locations A through E are in a nonforest condition). Starting at point 3 (the highest numbered regular point in the same condition as point 1) and at 0° azimuth, rotation at 60° intervals is performed to locate substitute locations for points 4 and 5.

SPATIAL STRUCTURE FOR FOREST POPULATIONS

Differences along the population boundaries where edge effect conditions exist were divided into two disjoint sets. The first set, called population edge effect, was when trees along the outside edge of the population differed from trees in the interior. We used the term edge effect between adjacent subpopulations when trees along the boundary of two adjacent subpopulations differed from interior trees in each subpopulation. For example, the transition zone between a hardwood and a softwood stand was an edge effect between adjacent subpopulations.

Population edge effect was important when sample points were substituted with points inside the population because the probabilities of selection for trees near the boundary change. However, the original probabilities of selection were used because the calculation of true probabilities of selection required additional measurements and was dependent on the location of point 1. Edge effect between adjacent subpopulations was less important because there can be offsetting estimation errors. For example, if points are moved out from a pine plantation at one location, it is possible that other points will be moved into a similar pine plantation at another location. Thus, it is difficult to determine situations where edge effect between adjacent population can cause a bias. However, rotation increases the variance of the parameter estimates because between-cluster variance is increased.

Some disturbances at the stand boundary that cause population edge effect are: Kudzu (*Pueraria lobata* Willd.(Onwi)) infestation, wind damage, reduced resource competition, increased light penetration, and differences in snow deposition.

Chen (1991) and Chen et al. (1992a, 1992b) studied the differences in microclimate between stand edge and stand interior for Douglas-fir (*Pseudotsuga meziesii* (Mirb.) Franco) forests in the Pacific Northwest region of the United States. They found significant differences in microclimates such as changes in average moisture content

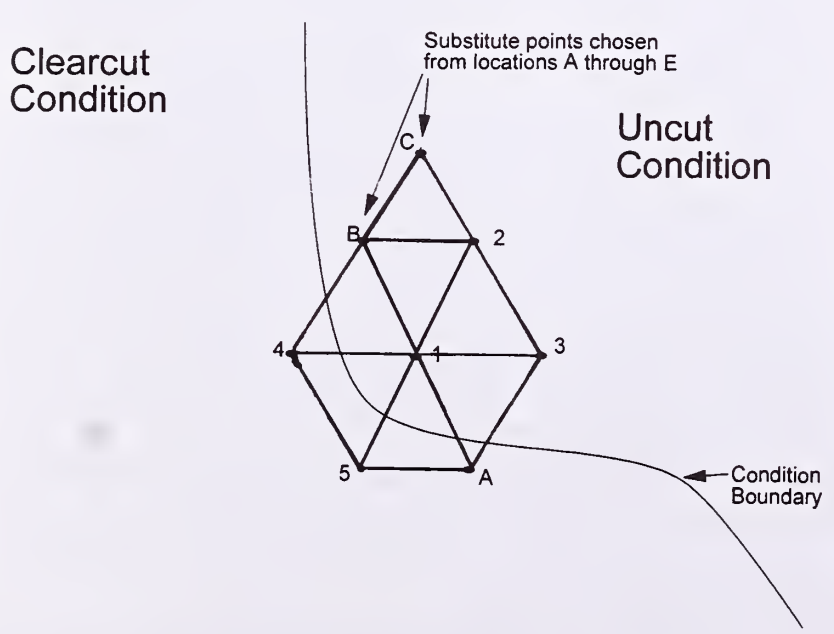


Figure 3. Example of plot substitution with lettered points A through E.

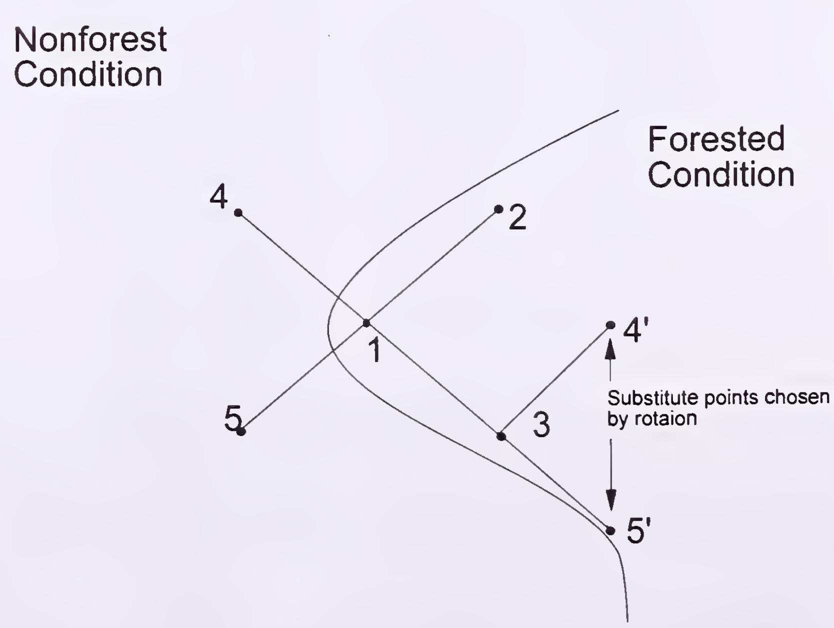


Figure 4. Example of rotation about regularly numbered points.

and temperature. The depth-of-edge influence ranged from 16 to 137 m for variables that were sensitive to the distance from the edge such as basal area, stem density, and canopy coverage.

DESCRIPTION OF DATA SETS

We were unable to find any satisfactory mapped data sets. The data set used consisted of a combination of distinctly different stem map data sets located near Sault Sainte Marie, Northern Ontario (i.e., a hardwood stand, softwood stand, and a pine plantation). This data set was extensively modified to show the effects of changing forest conditions. Ek (1969) provided a detailed description of the original data.

The combined data set was 48,204 trees on 67.423 ha. For each tree in the data set longitude, latitude, and diameter was available. One large population was constructed by combining 2 copies of each of the 3 data sets. Table 1 is a summary of the number of trees in each forest type.

For the combined data set, referred to as population 1, each forest type was homogeneous. This implies that edge effect between adjacent forest type and population edge effect were negligible or nonexistent. Also absent from the data set were disturbances such as unimproved roads, ponds, lakes, streams, and inclusions.

To study the effects of substituting points on estimation bias, more diverse populations were needed. Unfortunately, such mapped populations were unavailable. Thus, 14 modified populations were created from population 1 with varying degrees of boundary effects. Populations 2 through 15 were created by changing the diameters of trees within 15 m (49.2 ft) of the population edge. The range of diameter changes was from a 70 percent reduction to a 70 percent increase in increments of 10 percent. The 15 m distance was chosen because of the general hypothesis that edge effect will propagate into the stand at least 1 to 2 times the average tree height of the stand (personal communication with Wayne Shepperd). Chen et. al (1992b) studied a number of variables dependent on the distance from the edge of a clear-cut and found the depth-of-edge influence to range from 16 to 137 m. Therefore, the 15 m distance was probably a conservative estimate of the width of the edge effect for these types of stands. Slightly less than 9 percent of the population was affected by the diameter change.

Table 1. Description of the Northern Ontario data set.

Forest type	Number of trees	Mean diameter (in cm)
0 (Total)	48204	19.86
1 (Hardwood)	11006	25.65
2 (Softwood)	13622	15.87
3 (Pine Plan.)	23576	19.46

SIMULATION

In a simulation study, we documented that the practice of substituting points can cause substantial bias for basal area estimation. The Horvitz-Thompson estimator was used:

$$Y_{HT} = \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{n_j} \frac{Y_{ij}}{\pi_{ij}},$$

where:

m = 5 is the number of points in the cluster,
n_j = the number of trees sampled at point j,
Y_{ij} = the basal for tree i on point j,
and
π_{ij} = the probability of selection for a single tree.

The Intermountain FIA design, described earlier, was implemented once with points substituted (SUB) and once with points not substituted (NONSUB). The parameter of interest was basal area (BA) because it most closely related to tree volume. For the SUB and NONSUB designs, the selection probabilities were not corrected for trees along the population boundary. This was a reasonable assumption because no FIA unit adjusts selection probabilities for boundary trees and true probabilities would be difficult to compute for the substituted design. For comparison, an additional 5-point cluster design was implemented where all 5 points were selected randomly from the population and probabilities of selection where corrected for boundary trees. This design will be referred to as the RAND design.

The purpose of the RAND design was to help determine how many simulations were required before the estimates of BA stabilized, and to verify

that an unbiased estimator was attainable when selection probabilities were adjusted along the population edge. It is generally not appropriate to use the most efficient design to determine stability, but the bias of the SUB and NONSUB designs made it difficult to determine when stability was reached. Estimates from each of the 3 designs were generated for the entire population (forest type 0) and for each of the 3 forest types (forest types 1 through 3). Biases for the estimated basal area totals were expressed as a percentage of the actual population totals. For each iteration of the simulation, twenty 5-point cluster samples were drawn from the population. A total of 2,000 simulations were performed for each of the 15 populations described earlier.

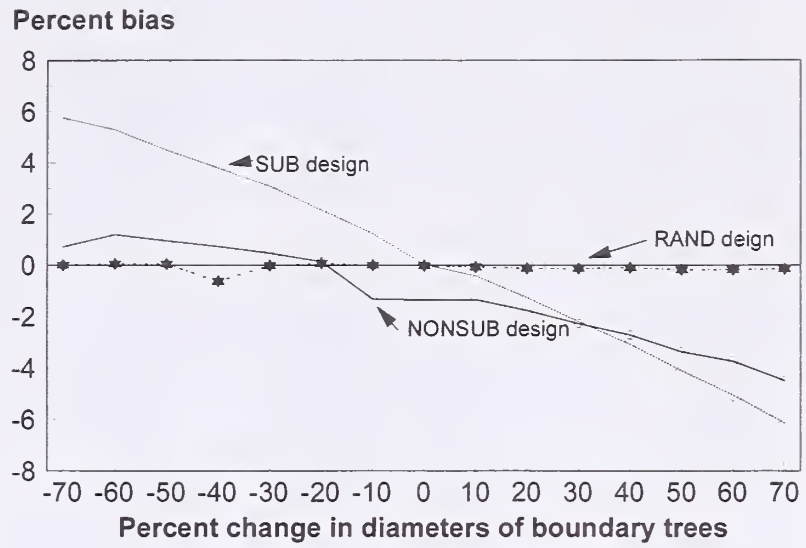


Figure 5. Bias expressed as a percentage of population basal area versus percent change in diameter of the boundary trees (9 percent of all trees subject to diameter change).

RESULTS AND DISCUSSION

A bias was documented for populations where edge effect was present. Figure 5 shows the bias, expressed as a percentage of the total basal area, versus the percent change in diameters along the boundary for both the SUB, NONSUB and RAND designs for forest type 0. The figure shows that the bias versus percent change along the edge of the population had a strong linear relationship. Fitting the data gives the following models:

$$B(NONSUB) = -1.144 - 0.041*PERCENT \text{ with } R^2 = 0.958,$$

$$B(SUB) = 0.241 - 0.086*PERCENT \text{ with } R^2 = 0.996,$$

and

$$B(RAND) = -0.052 + 0.002*PERCENT \text{ with } R^2 = .072.$$

where B(NONSUB), B(SUB), and B(RAND) were the bias for the NONSUB, SUB, and RAND designs, respectively. PERCENT was the percent change of the diameters along the boundary (-70 to 70). The population bias ranged from -6.2 to 5.8 percent for the SUB design, from -4.5 to 1.2 percent for the NONSUB design, and from -0.2 to 0.08 percent for the RAND design. Similar results were found for forest types 1 through 3. The model coefficients for the SUB and NONSUB designs are listed in table 2. The models for forest types 0 through 3 showed a strong linear trend indicating that as the diameters of trees along the population edge increased, the SUB design gave an underestimate of the basal area. When the average diameter along the population edge decreased, the SUB design showed an overestimate of the basal area. One source of the bias was the changing selection probabilities for individual trees along the population boundary. When points were moved away from the population boundary the probability of selection for boundary trees was lowered. Thus, when population edge effect was present, trees in different conditions were given incorrect selection

Table 2. Model coefficients for bias versus percent change in diameter of edge effect trees.

	Substituted			Nonsubstituted		
Type	α	β	R^2	α	β	R^2
0 (Total)	0.241	-0.086	0.996	-1.144	-0.041	0.958
1 (Hardwood)	1.434	-0.074	0.979	-0.671	-0.050	0.952
2 (Softwood)	0.796	-0.062	0.974	-2.384	-0.036	0.927
3 (Pine plan)	0.877	-0.112	0.998	-0.666	-0.043	0.968

probabilities, which increased estimator bias. The second source of bias was caused by the 5-point cluster. The fixed location of points 2 through 5 caused trees in the population interior to be sampled with a higher intensity than trees along the population boundary.

The NONSUB design also showed a tendency for bias to increase as population edge effect increases. This was due to differing sampling intensities at the population boundary. When the tree diameters along the edge increased, the difference between the unadjusted and adjusted selection probabilities and the number of trees that required adjustment of their selection probabilities increased. The differences between the actual and assumed selection probabilities were less for the NONSUB design. The RAND design produced unbiased estimates because the correct selection probabilities were used and the random placement of all 5-points ensured the interior and the boundary of the population were sampled with the same intensity. All estimates of bias fluctuate about zero, but with a slight increase in bias, this could be caused by the random number generator used. The RAND design showed that the data to compute the actual probabilities of selection should be collected in the field if possible.

Closer inspection of figure 5 shows that when no population edge effect was present (percent change = 0) the substituted design was less biased than the NONSUB design. The bias for basal area, as depicted in the graph, is -1.35 percent for the NONSUB design and 0.04 percent for the SUB design. In this case, plot substitution reduced the bias of the estimates because for trees along the boundary the actual selection probability was smaller than the theoretical one, thus an underestimate of the population total is expected.

Table 3. Average increase in efficiency for the nonsubstituted design expressed as a percentage of the SUB design and given by 100 * STDEV(NONSUB)/STDEV(SUB).

Forest type	Average efficiency
0	94.7
1	93.9
2	93.4
3	94.5

To determine if point substitution had an affect on the efficiency of basal area estimation, the simulation standard deviations were compared for populations 2 through 15. The average percentage change in simulation standard deviation was calculated by taking the average of:

EFFICIENCY = 100*STDEV(NONSUB)/STDEV(SUB)

for data sets 2 through 15 using all forest types, where STDEV(NONSUB) and STDEV(SUB) are the simulation standard deviation for the NONSUB and SUB designs, respectively (table 3). In all cases, the NONSUB design was more efficient than the SUB design. The increase in efficiency for the NONSUB design was because more heterogeneous samples were obtained in each of the 5-point clusters. When points were substituted, the 5-point clusters were more homogeneous; therefore, the differences between clusters was greater, which decreases efficiency.

RECOMMENDATIONS

The bias of a substituted point design can be minimal or substantial depending on the population characteristics. In certain regions of the United States, up to 30 percent of plots have been substituted to keep all plots in the same forest type. Biases of more than 6 percent were shown to commonly exist. The bias should be more serious for small stands with considerable population edge effect. Even for the NONSUB design, it is important that the correct selection probabilities are computed or at least approximated well. In addition to decreased estimation bias, the NONSUB design was more efficient then the SUB design. If a change was made to a design where points were not substituted, no concerns about the efficiency of the estimates for the current sample grid would be raised. Therefore, a nonsubstituted sampling design should be used to avoid estimation bias and increase efficiency.

LITERATURE CITED

Chen, J. Q. 1991. Edge effects: Micro-climatic pattern and biological responses in old-growth Douglas-fir forests. Ph.D. Thesis, U. of Wash. Seattle, WA.

Chen, J. Q., Franklin, J. F. and Spiess, T. A. 1992a. Contrasting microclimates among clear-cut, edge, and interior of old-growth Douglas fir forests. Ecol. Appl. 63(3-4): 219-237.

Chen, J. Q., Franklin, J. F. and Spiess, T. A. 1992b. Vegetation responses to edge environments in old-growth Douglas-fir forests. Ecol. Appl. 2: 387-386.

Ek, A. R. 1969. Stem map data for three forest stands in Northern Ontario. Information Report 0-X-113. For. Res. Lab. Ontario Region. Sault Ste. Marie, Ontario. Can. Forestry Service. Dept. of Fisheries and Forestry.

Lentz, G. H. 1932. Field manual for the Forest Survey in the bottomland hardwoods of the Mississippi Delta. Field Manual for the Forest Survey. Unpublished manuscript. USDA Forest Serv. Southern For. Exp. Sta. New Orleans, La. 30p.

Spada, B. 1960. A test of several designs for sampling an acre to obtain forest survey volume and area statistics and area condition classification data. Internal document PNW Station.

USDA Forest Service Intermountain Forest Experiment Station. Idaho Forest Survey Field Procedures 1990-1991. IM Res. Sta. Ogden, UT. 181 p.

Rocky Mountains

Southwest

Great Plains

U.S. Department of Agriculture
Forest Service

Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of seven regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico
Flagstaff, Arizona
Fort Collins, Colorado*
Laramie, Wyoming
Lincoln, Nebraska
Rapid City, South Dakota

*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526